TECHNICAL MEMORANDUM

Draft Radiological Data Evaluation Plan, Former Hunters Point Naval Shipyard (HPNS)

PREPARED FOR:	HPNS Radiological Tiger Team
PREPARED BY:	CH2M HILL, Inc. (CH2M), Cabrera Services, Perma-Fix Environmental Services, Inc.
DATE:	January 2017
CONTRACT:	Comprehensive Long-term Environmental Action—Navy (CLEAN) Contract N62470-16-D-9000, Contract Task Order FZ12

1 Introduction

This Technical Memorandum presents the overall plan and data quality objectives (DQOs) for evaluation of historical soil data and confirmation sampling at the former Hunters Point Naval Shipyard (HPNS), San Francisco, California. The iterative DQO process will be actively utilized during this project to optimize the approach (e.g., adding or combining tests and decision rules based on investigator and stakeholder input) throughout the entire investigation.

2 Data Evaluation Plan

A phased data evaluation approach is planned with parallel work paths (**Figure 1**) to evaluate radiological data and conduct confirmation sampling. This approach is responsive to stakeholder concerns (**Figure 2**) and will provide multiple lines of evidence from a consortium of health physics and human health risk assessment professionals for decision-making. The plan for Phase 1 is to develop a soil database with available data, conduct soil data evaluation to identify anomalies indicating potential data falsification while validating sources of usable data, and identify data gaps for Phase 2 evaluation. Although Phase 2 is still under development, the plan is to identify additional data needs, conduct detailed review to further evaluate anomalous data identified during Phase 1 and on additional data, and design and conduct confirmation sampling. The key steps of the phased data evaluation approach are summarized below.

Phase 1 Evaluation:

- Compare Naval Installation Restoration Information Solution (NIRIS) soil data to the Tetra Tech EC soil database and available final status survey reports to assess completeness and accuracy of the data sets. A statistically determined number of reports and individual data results from written reports (RACRs/SUPRs) will be compared with electronic results of NIRIS and Tetra Tech EC databases to identify potential manipulation. This analysis is biased toward reports from sites with areas with the greatest exceedances of cleanup levels prior to remediation being selected more often than reports from sites with less exceedances prior to remediation. Data will also be compared to laboratory reports once they are obtained. It should be noted that RASO reviewed original laboratory data reports so falsification of the laboratory results database, if it occurred, is not likely to have affected decision-making.
- Organize and separate soil data sets by survey units/trench units, work areas, parcel, and survey type to include: scoping, characterization, post-remediation, biased, final status, and use of excavated soil for backfill.

- Use the assembled soil data sets to run statistical evaluation/presentation within and among survey unit data sets as well as larger data sets such as work areas and parcels in order to identify unexpected or unusual data.
    - Evaluate summary statistics and normal probability plots for all soil datasets to identify outliers as unusual sample results and identify multiple or unexpected distributions of data as unusual datasets.
    - Investigate equilibrium status of radium-226 and thorium-232 decay series using the Peacock test to identify unusual sample results and unusual datasets.
    - Use Analysis of Variance (ANOVA), Runs Tests, Benford's Law, and Peacock Test for additional investigation of unusual or unexpected results.
    - Excel or Pro-Upper Confidence Level (UCL) statistical analysis will provide most of the analytical capability. Once the data is segregated and checked, the evaluation will be automated (i.e., macro-driven) as much as possible.
- Identify missing data or incomplete datasets.
- Separate and report data that requires more in depth investigation in Phase 2.

Screening release criteria used in statistical tests to evaluate falsification are specified in Table 1 of the Final Basewide Radiological Removal Action, Action Memorandum Revision 2006, April 21, 2006.

Phase 2 Evaluation (still under development):

- Review available records for the entire history of Tetra Tech EC radiological work that produced soil data to support property transfer decisions, including areas already transferred.
- Perform in-depth review of anomalous data identified in Phase 1 to include: additional statistical analyses, hard copy site records review (i.e., review of physical and electronic sample documentation), visual evaluation of archived samples, and collection of confirmatory samples to provide multiple lines of objective evidence.
- Evaluate all data for significance in decision-making.
- Identify areas with the highest concentrations of radionuclides based on pre-remediation data. These areas may be targeted for detailed review of existing data and/or further field data collection efforts.
- Prepare work plan(s) and Uniform Federal Policy–Sampling and Analysis Plan (UFP-SAP) for sample collection that focus on results from investigation of Phases 1 and 2 results and previous allegations and concerns. Ensure review and input from stakeholders.
- Collect data to replace manipulated data, fill missing data or incomplete datasets. Perform confirmatory sampling field work. Provide schedule notification to stakeholders for observation or independent sampling and analyses.
- Combine Phases 1 and 2 results into a Final Report. Ensure review and input from stakeholders.

3 Data Quality Objectives

The DQOs for this evaluation are presented below. The DQOs are intended to be part of a living data evaluation process and updated as additional information is obtained and through the Environmental Protection Agency (EPA) Triad Approach (see Step 7).

Step 1 – Identify the Problem

The Navy goal is to release real property from radiological controls. The real property includes seven parcels at the former Hunters Point Naval Shipyard. The existing data was collected using approved plans based on a foundation of HSA information and risk-based criteria documented in the Record of Decision (ROD). There have since been claims of data manipulation and falsification of records. The objectives are as follows:

- ~~The primary objective is to e~~Evaluate ~~and communicate~~ the validity of the existing (or previous) radiological data to see if it can be used to support decisions regarding transfer of property at HPNS.
- ~~The secondary objective is to c~~Confirm or corroborate data evaluation findings and/or prior sampling results with additional soil samples (e.g., confirmation samples) if, after the validation process, the stakeholders believe additional data is necessary.
- ~~If the data evaluation findings identify new anomalous samples, new/supplemental sampling will be conducted and the results evaluated to determine if the new sampling results can be used to support the decision to transfer property.~~
- Communicate~~d~~ the validity of the existing and new (if any) radiological data to the public.

Identify Members of the Planning Team– Navy Base Realignment And Closure (BRAC), Naval Facilities Engineering Command Southwest (NAVFAC SWDIV), Naval Sea Systems Command Detachment, Radiological Affairs Support Office (NAVSEADET RASO), EPA, California Department of Toxic Substances Control (DTSC), City of San Francisco, CH2M, Cabrera Services, Perma-Fix Environmental Services, Inc., and SC&A Environmental Services and Consulting.

Identify Primary Decision Maker – Navy

Identify Decision-Making Method – Perform data quality assessment of verified and validated data provided by the Navy to identify individual samples or datasets with anomalous results that could indicate falsified data (Phase 1). Perform additional investigation (either additional assessment, data review, field measurement, or sample collection and analysis) on the anomalous data (Phase 2). Summarize the assessment results and provide recommendations on whether the existing data should be used to support property transfer decisions.

Specify Available Resources and Relevant Deadlines – The labor and technical resources are provided by CH2M and its subcontractors. Available resources include lab analytical soil data (approximately 70,000 records) and electronic reports (approximately 475 reports in pdf file format) from NIRIS data and lab analytical data for soil provided by Tetra Tech EC (approximately 2 million records). Additional qualitative scan and static gross gamma data are available and may be utilized in the Phase 2 investigation.

Commented [HK1]: Response to Nina's addition: This addition is not part of the problem statement and is one of several lines of evidence throughout Phase 2 where appropriate, see DTSC RTC #2.

Step 2 – Identify the Decision

The decision to be made is whether the data to support transfer of real property has been manipulated. ~~While the secondary decision is whether supplemental sampling will be sufficient to support the transfer of real property if the data has been manipulated.~~

If the Team identifies that the data used to support a decision to transfer real property has been manipulated, then the data will not be used for property transfer decisions and supplemental data will be collected to replace manipulated data; otherwise the Team will conclude that the data can be used to support transfer of the real property. Questions to be answered for original and supplemental data are:

- Any evidence levels of radioactivity exceed the release criterion?

Commented [HK2]: Response to Nina's addition: Sampling is one of several lines of evidence to support decisions for property transfer.

- Any evidence of unusual data (outliers, disequilibrium, unexplained shifts in concentration) that requires additional investigation to support property transfer decisions?
- Any evidence of line-by-line data manipulation (e.g., changing an individual dataset)?
- Any evidence that soil samples were not representative of the matrix (e.g., not collected from the survey unit soil matrix)?
- Confirmation samples contradict existing data conclusions?

For Phase 2, if a decision cannot be made on whether the data has been manipulated, new decision rules will need to be developed. Specific falsification allegations will be evaluated by reviewing dates of sample collection, chains of custodies, log book entries, inspection of sample archives, and draft and final reports as available and decision rules will be developed for these proposed activities.

Step 3 – Identify Inputs to the Decision

The initial scope of inputs is all soil data and the documents prepared and approved for all parcels and survey units. Inputs for the evaluation also include a list of data falsification claims. Complete definition of data inputs will be clarified as the database is collected, prepared, checked for accuracy, and the evaluations are performed. Potential sources of inputs that may have significant information, but a complete set have not been available to date are: field logs, sample logs, and sample chain of custodies.

For Phase 1, the following tests will be performed to evaluate data:

- Summary Statistics (mean, median, standard deviation, count)
- Histograms
- Normal Probability Plots
- Posting Plots
- Time Series Plots
- Kruskal-Wallis Test One-way Analysis of Variance (ANOVA)
- Two-Dimensional Paired Kolmogorov-Smirnov (Peacock) Test
- Repeated Numbers Test
- Benford's Law Test

For Phase 2, the following lines of investigation based on Phase 1 input will provide input to the data evaluation (Note: additional lines of evidence may be added as the project develops):

- Additional statistical analyses based on Phase 1 results (e.g., different groupings of contiguous survey units, different statistical tests to answer specific questions about the data).
- Physical review of archived samples to inspect matrices (physical appearance and geology of samples) for consistency within and among survey units.
- Review sample documentation through sample logs, field logs, and chain of custody.
- Comparison of lab data to final report data.
- Test pitting or confirmation sample, or supplemental sample collection(s) as either directed by Phase 1 or Phase 2 investigation results or stakeholder recommendations and concerns.

Step 4 – Identify Boundaries for the Study

Temporal boundary runs through the end of the contract task order, currently 11/18/18 (allowing two months for project closeout, technical completion date is 9/18/18). The physical boundaries are Parcels B, C, D, E, E-2, F, G, UC-1, UC-2, and UC-3 at HPNS. The data boundaries are limited to available soil data during Phase 1 that includes lab analytical soil data and electronic reports from NIRIS and lab analytical data for soil provided by Tetra Tech EC. In Phase 2, the data boundaries are limited to available data provided by Phase 1 and data obtained in hard copy or from Tetra Tech EC based on gaps identified and

data requests. The survey boundaries are the complete set of final status survey soil analytical laboratory data based on the claims of data falsification for soil sample data only. Additional boundaries will be determined for Phase 2.

Step 5 – Develop a Decision Rule – To be refined during both Phase 1 and Phase 2

If the Team identifies that the final status survey soil sample data collected by Tetra Tech EC for transfer of real property from Parcels B, C, D, E, E-2, F, G, UC-1, UC-2, and UC-3 at HPNS has been manipulated, then the data will not be used for property transfer decisions; otherwise, the Team will conclude that the data can be used to support decisions for transfer of the real property. If a decision cannot be made on whether the data has been manipulated, new decision rules will need to be developed.

Phase 1 Decision Rules:

- Summary Statistics
  - If any analytical result from a final status survey is more than three standard deviations away from the mean for that data set, that result will be flagged for investigation in Phase 2.
  - If any analytical result from a final status survey exceeds the Derived Concentration Guideline Level- Wilcoxon Rank Sum test ($DCGL_W$), that result will be flagged for investigation in Phase 2.
  - If any result plus the reported total uncertainty is less than zero (e.g., a negative number), that result will be flagged for investigation in Phase 2.
- Histograms
  - Compare onsite lab results with offsite lab results for each radionuclide for individual survey units (when data are available).
  - Compare final status survey systematic results with pre-remediation systematic results for each radionuclide for individual survey units (when data are available).
  - Compare individual survey unit data with all available data from a specific site for each radionuclide.
  - Flag entire data set for investigation in Phase 2 when significant differences are identified using professional judgement.
- Normal Probability Plots (Cumulative Frequency Distribution, Quantile Quantile [QQ] Plot)
  - Compare onsite lab results with offsite lab results for each radionuclide for individual survey units (when data are available).
  - Compare final status survey systematic results with pre-remediation systematic results for each radionuclide for individual survey units (when data are available).
  - Straight-line plot indicates normal distribution consistent with expectations for background, ~~flag individual points off the line for investigation in Phase 2, flag final status survey data sets significantly different from pre-remediation data sets for investigation in Phase 2,~~ flag final survey data sets with significantly non-linear plots (curves and changes in slope) for investigation in Phase 2.
- Posting Plots
  - Plot final status survey systematic sample locations for each survey unit.
  - If data points are missing or do not follow a systematic pattern, flag the data set for investigation in Phase 2.
- Time Series Plots

  - Plot all analytical results for a specific radionuclide from a site or area in order by sample collection date.
  - Plot the mean of all results for a specific radionuclide from samples collected on a specific day.
  - Plot the precision of all results for a specific radionuclide from samples collected on a specific day.
  - Identify results that show visual trends to be investigated in Phase 2.
- Kolmogorov-Smirnov (K-S) Test
  - Compare data from one survey unit against data from all other survey units or data from one day against data from all other days based on rankings.
  - Use the time series plots to identify unusual data not identified by the K-S Test results.
  - Data collected in a unit where the average result for that unit or the variance for that unit is significantly different from data collected in nearby units will be investigated in Phase 2.
  - Data collected on a day~~s~~ where ~~individual sample results or~~ the average result for that day ~~are~~ or the variance for that day is significantly different from data collected on previous or following days will be investigated in Phase 2.
- Kruskal-Wallis One-Way Analysis of Variance (ANOVA)
  - Compare radionuclide-specific results from multiple survey units to determine if one or more survey units are from different distributions.
  - Compare radionuclide-specific results from multiple days of sample collection activities to determine if one or more day's samples were collected are from different distributions.
  - If the calculated p-value from the Kruskal-Wallis test is less than the critical value from the Chi-square distribution, there are one or more survey units with data from a different distribution. Use the Kolmogorov-Smirnov Test to identify the data sets from different distributions. Flag data sets for survey units or sample collection days from different distributions for investigation in Phase 2.
- Two-Dimensional Paired Kolmogorov-Smirnov (Peacock) Test
  - Compare radium-226 offsite laboratory results for a data set with lead-214 offsite laboratory results for the same data set to evaluate equilibrium conditions.
  - Compare thorium-232 offsite laboratory results for a data set with lead-212 offsite laboratory results for the same data set to evaluate equilibrium conditions.
  - Compare thorium-232 offsite laboratory results for a data set with bismuth-212 offsite laboratory results for the same data set to evaluate equilibrium conditions.
  - If the p-value is below 0.05 there is a significant difference in the distribution of the parent and the distribution of the progeny, and the data set will be investigated in Phase 2. The p value of 0.05 is standard and 100 (1-p)% gives a confidence level of 95%.
- Repeated Numbers/Runs Test
  - Identify days where more than two results are identical for a specific radionuclide at a specific site.
  - Identify survey units from a specific site where more than two final status survey results for a specific radionuclide are identical.

  - If more than ~~25% of~~two results for a specific radionuclide are duplicated between two data sets, both data sets will be investigated in Phase 2.

- Benford's Law Test
  - ~~The only digits that are being analyzed in the Benford's Test are the first or the first and the second non-zero digits.~~
  - Count the number of times a digit (i.e., 1, 2, 3, 4, 5, 6, 7, 8, or 9) appears as the first number in each result for a data set.
  - If the number of times each digit appears is inconsistent with Benford's Law, that data set will be investigated in Phase 2.

- General Logic
  - If more than three rounds of remediation were performed, that survey unit will be flagged for investigation in Phase 2 and confirmatory samples will be recommended in that survey unit based on an approved plan.
  - If the preparation date, count date, or reporting date for any sample occurs before the sample collection date, that sample will be evaluated in Phase 2.
  - If a sample listed in the NIRIS database does not have a corresponding sample in the Tetra Tech EC database, that sample will be investigated in Phase 2.
  - If a record in the NIRIS database is different from the corresponding record in the Tetra Tech EC database, that record will be investigated in Phase 2.

Phase 2 Decision Rules:

- If the data in the Tetra Tech EC database are statistically evaluated against the hard copy laboratory reports using MIL-STD-~~1916~~19074, "DoD Preferred Methods for Acceptance of Product" (available online at [HYPERLINK "http://assist.daps.dla.mil/"]), or ANSI/ASQ Z1.4, "Sampling Procedures and Tables for Inspection by Attributes" (available online at [HYPERLINK "http://www.asq.org/"])~~Sampling Procedures and Tables for Inspection by Attributes~~ and the data match to the 95% confidence level, then the data are consistent, and may not represent substituted data pending the results of other lines of investigation.
- If data appear in the NIRIS database but no corresponding data appear in the Tetra Tech EC database, that data may represent substituted data and will be require a physical review of available sample documentation evidence to be assembled for data validation.
- If data appear in hard copy reports and are not reported in the electronic database or vice-versa, the data may be falsified.
- If inconsistencies are found during project sample and survey paperwork review (sample logs, field logs, and chain of custodies) of Phase 1 flagged data (e.g., improbable timing of events), then the data may have been falsified and the multiple lines of objective evidence will be documented.
  - If samples were collected or analyzed after remediation was complete at site areas (e.g., FSS samples), then the data may have been falsified.
  - If sample results had minimum detectable concentrations that exceeded the DCGL and the site was declared to have met the clean-up criteria and was released, then the site may actually exceed the clean-up criteria and more samples should be collected at that site area.
  - If biased samples were collected before the gamma survey was performed, then the data may be falsified.

  - If project logbooks are available and are inconsistent with sample chain of custody, then the data may have been falsified.
  - If sample chain of custody and/or logbooks (if available) show samples were sent to a laboratory different than the laboratory that reported the results, then the data may be falsified.
  - If the result from one or more sample analyses from a pre-remediation, characterization, or remedial action support survey are identified as statistical outliers in Phase 1 and no remediation was performed at that sample location and the final status survey results do not indicate corresponding outliers, the final status survey data for that survey unit will be flagged as questionable.
  - If the results from an individual sample or survey unit are flagged as questionable from two or more data evaluation sources, the final status survey data for that sample or survey unit will be rejected (data validation R flag) and will not be used to support decisions regarding property transfer.
- If indicated by the types, locations, quantities of data flagged by the Phase 1 evaluation, then additional statistical testing (e.g., different groupings of contiguous survey units, or a different statistical test) may be added to the DQOs and used to further investigate similar survey units or matrices.
- If survey unit or parcel location test pitting or visual evaluation of archived samples based on investigation of flagged Phase 1 data show inconsistencies with the sample and survey unit matrix, then the data may have been falsified and the multiple line of objective evidence will be documented.
- When insufficient information or data are available to complete investigation of area(s), parcels or survey units, then those areas, parcels or survey units will have insufficient data to support property transfer decisions.
- If confirmatory samples find activity above the release criteria or with naturally occurring decay series concentrations that vary significantly from the previous site sample results, then the data may have been falsified and the multiple line of objective evidence will be documented.
  - Sampling will be focused toward historically high risk areas, EPA recommendations and concerns, the approved conceptual site model, and areas with known or suspected data manipulation.

Step 6 – Specify Limits on Decision Errors

The null hypothesis is that there are data that indicate one or more parcels, sites, areas, or survey units have data that has been manipulated by falsification. By reviewing 100% of the data we minimize the Type I error (deciding a parcel is acceptable for release when it is not).

The multiple lines of evidence approach to document potential falsification of data limits the Type II error and mitigates the possibility of deciding a parcel or survey unit is not acceptable when it really is.

Step 7 – Optimize the Design

During the course of the investigation the Team will refine the evaluation methodology for additional uses of currently described testing, other tests or graphs, and additional sources or types of information in order to optimize the evaluation of data manipulation. The EPA Triad Approach will drive the implementation and optimization of project methodology to manage decision uncertainty, thus increasing confidence that project decisions are made correctly and cost-effectively. The project optimization will occur through:

  - Systematic project planning (routine planning meetings are scheduled throughout project life),

- Dynamic work strategies (phased approach implemented by separate teams coordinated through leadership and planning), and
- Real-time assessment and sharing of data (dynamic feedback of results between and among Phases 1 and 2 teams and stakeholders).

~~Schedule~~

Commented [HK3]: Response to Nina's addition: A project schedule will be developed, reviewed, and approved by the Tiger Team separately.

Figures